

# Maximum weighted independent sets with a budget

Tushar Kalra<sup>1</sup>, Rogers Mathew<sup>1</sup>, Sudebkumar Prasant Pal<sup>1</sup>, and Vijay Pandey<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering,  
Indian Institute of Technology  
Kharagpur 721302, West Bengal, India.

{tushar1nitjkalra, rogersmathew, sudebkumar, vijayiitkgp13}@gmail.com

## Abstract

Given a graph  $G$ , a non-negative integer  $k$ , and a weight function that maps each vertex in  $G$  to a positive real number, the *Maximum Weighted Budgeted Independent Set (MWBIS) problem* is about finding a maximum weighted independent set in  $G$  of cardinality at most  $k$ . A special case of MWBIS, when the weight assigned to each vertex is equal to its degree in  $G$ , is called the *Maximum Independent Vertex Coverage (MIVC) problem*. In other words, the MIVC problem is about finding an independent set of cardinality at most  $k$  with maximum coverage.

Since it is a generalization of the well-known Maximum Weighted Independent Set (MWIS) problem, MWBIS too does not have any constant factor polynomial time approximation algorithm assuming  $P \neq NP$ . In this paper, we study MWBIS in the context of bipartite graphs. We show that, unlike MWIS, the MIVC (and thereby the MWBIS) problem in bipartite graphs is NP-hard. Then, we show that the MWBIS problem admits a  $\frac{1}{2}$ -factor approximation algorithm in the class of bipartite graphs, which matches the integrality gap of a natural LP relaxation.

**Keywords:** Independent set, partial vertex cover, coverage, approximation algorithm, NP-hard, inapproximability.

## 1 Introduction

### 1.1 Problem definition

Let  $G$  be a graph and let  $w : V(G) \rightarrow \mathbb{R}^+$  be a function that assigns positive real numbers as weights to the vertices of  $G$ . Under this assignment of weights, for any set  $S \subseteq V(G)$ , we define the weight of  $S$ , denoted by  $w(S)$ , as the sum of the weights of the vertices in  $S$ . The famous *Maximum Weighted Independent Set (MWIS) problem* is about finding an independent set of vertices in  $G$  that has the highest weight amongst all the independent sets in  $G$ . In this paper, we study a budgeted version of the well-studied MWIS problem, namely *Maximum Weighted Budgeted Independent Set (MWBIS) problem*.

**Definition 1.** Given a graph  $G$ , a weight function  $w : V(G) \rightarrow \mathbb{R}^+$ , and a positive integer  $k$ , the *MWBIS problem* is about finding an independent set of size at most  $k$  in  $G$  that has the highest weight amongst all independent sets of size at most  $k$  in  $G$ .

### 1.2 Related work

A more general problem (also known by the same name, MWBIS,) was introduced and studied in the context of special graphs like trees, forests, cycle graphs, interval graphs,

and planar graphs in [5] where each vertex in the given graph  $G$  has a cost associated with it and the problem is about finding an independent set of total cost at most  $C$  (, where  $C$  is a part of the input,) in  $G$  that has the highest weight amongst all such independent sets. Apart from this work, to the best of our knowledge, not much is known about MWBIS.

Given a graph  $G$ , we know that the *Vertex Cover (VC) problem* is about finding the minimum number of vertices that cover all the edges of  $G$ . Several variants of the VC problem has been studied in the literature. We discuss about a couple of them here. For a positive integer  $t$ , the *Partial Vertex Cover (PVC) problem* is about finding the minimum number of vertices that cover at least  $t$  distinct edges of  $G$ . In the year 1998, Burroughs and Bshouty introduced and studied the problem of partial vertex cover [9]. In this paper, the authors gave a 2-factor approximation algorithm by rounding fractional optimal solutions given by an LP relaxation of the problem. Bar-Yehuda in [6] came up with another 2-approximation algorithm that relied on the beautiful ‘local ratio’ method. A primal-dual algorithm achieving the same approximation factor was given in [11]. In [10], it was shown that the PVC problem on bipartite graphs is NP-hard.

Another popular variant of the VC problem is the *Maximum Vertex Coverage (MVC)* problem. Given a graph  $G$  and a positive integer  $k$ , the *MVC problem* is about finding  $k$  vertices that maximize the number of distinct edges covered by them in  $G$ . Ageev and Sviridenko in [1] gave a  $3/4$ -approximation algorithm for the MVC problem. An approximation algorithm, that uses a semidefinite programming technique, based on a parameter whose factor of approximation is better than  $3/4$  when the parameter is sufficiently large was shown in [14]. Apollonio and Simeone in [3] proved that the MVC problem on bipartite graphs is NP-hard. The same authors in [2] gave a  $4/5$  factor approximation algorithm for MVC on bipartite graphs that exploited the structure of the fractional optimal solutions of a linear programming formulation for the problem. The authors of [10] improved this result to obtain an  $8/9$  factor approximation algorithm for MVC on bipartite graphs.

### 1.3 MIVC problem - a special case of MWBIS

Let us come back to our problem - the MWBIS problem. In this problem, what happens if the weight function given maps each vertex to its degree in  $G$ ? That is, let  $w(v) = \deg(v)$ ,  $\forall v \in V(G)$ . We call this the *Maximum Independent Vertex Coverage (MIVC) problem*.

**Definition 2.** Given a graph  $G$ , a weight function  $w : V(G) \rightarrow \mathbb{R}^+$  defined as, for each vertex  $v \in V(G)$ ,  $w(v) = \deg(v)$ , and a positive integer  $k$ , the MIVC problem is about finding an independent set of size  $k$  in  $G$  that has the highest weight amongst all independent sets of size at most  $k$ .

Observe that the MIVC problem, as its name suggests, can also be seen as a variant of the MVC problem where the  $k$  vertices that we choose need to be an independent set. This observation gives us the following alternate definition.

**Definition 3.** Given a graph  $G$  and a positive integer  $k$ , the *MIVC problem* is about finding at most  $k$  independent vertices that maximize the number of edges covered by them in  $G$ .

### 1.4 IP formulation of MWBIS

Let  $G$  be a graph on  $n$  vertices, where  $V(G) = \{v_1, \dots, v_n\}$ , and let  $w : V(G) \rightarrow \mathbb{R}^+$  be a weight function given. Let  $k$  be a positive integer given. Let  $\mathcal{C}$  denote the set of all

maximal cliques in  $G$ . In order to formulate MWBIS on  $G$  as an integer program, let us assign a variable  $x_i$  for each  $v_i \in V(G)$ , which is allowed 0/1 values. The variable  $x_i$  will be set to 1 if and only if the vertex  $v_i$  is picked in the independent set.

$$\begin{aligned}
& \text{Maximize} \sum_{i \in [n]} w(v_i) \cdot x_i \\
& \text{s.t.} \sum_{i=1}^n x_i \leq k \\
& \quad \sum_{i: v_i \in C} x_i \leq 1, \quad \forall C \in \mathcal{C} \\
& \quad x_i \in \{0, 1\}, \quad \forall i \in [n].
\end{aligned} \tag{1}$$

The constraint  $\sum_{i=1}^n x_i \leq k$  ensures that not more than  $k$  vertices are picked. The constraint  $\sum_{i: v_i \in C} x_i \leq 1$ , for each maximal clique  $C \in \mathcal{C}$ , ensures that the set of selected vertices is an independent set. Replacing  $w(v_i)$  with  $\deg(v_i)$  in the objective function of (1) yields the IP formulation of MIVC for  $G$ .

Later in Section 2.3, we consider the LP-relaxation of this integer program. We show an instance of the MIVC problem (a bipartite graph) to illustrate the fact that the integrality gap of this LP relaxation is not ‘good’. This rules out the possibilities of using this relaxation to obtain good LP-based approximation algorithms for the MIVC problem (and thereby for the MWBIS problem) in bipartite graphs.

## 1.5 Hardness results

It is well-known that the MWIS problem cannot be approximated to a constant factor in polynomial time, unless  $P=NP$ . For every sufficiently large  $\Delta$ , there is no  $\Omega(\frac{\log^2 \Delta}{\Delta})$ -factor polynomial time approximation algorithm for MWIS in a degree- $\Delta$  bounded graph, assuming the unique games conjecture and  $P \neq NP$  [4]. As MWBIS is a generalization of MWIS, these results hold true for the MWBIS problem too. But, what about the special case of MWBIS - the MIVC problem? Below, we show that it is hard to approximate MIVC within a certain factor. We prove this by giving an approximation factor preserving reduction from another problem, namely 3-Maximum Independent Set (3-MIS) problem. The 3-MIS problem is about finding an independent set of maximum cardinality in a given 3-regular graph.

**Theorem 1.** *There is no  $(\frac{139}{140} + \epsilon)$ -factor approximation algorithm for the MIVC problem, assuming  $P \neq NP$ , where  $\epsilon > 0$ .*

*Proof.* It was shown in (Statement (v) of Theorem 1 in) [7] that there is no  $(\frac{139}{140} + \epsilon)$ -factor approximation algorithm for the 3-MIS problem, assuming  $P \neq NP$ , where  $\epsilon > 0$ . If there existed a polynomial time  $f$ -factor approximation algorithm  $A(G, k)$  for the MIVC problem, where  $f > 139/140$ , then we could use it as an  $f$ -factor approximation algorithm for the 3-MIS problem as follows. Given a positive integer  $k$  and a 3-regular graph  $G$  on  $n$  vertices as input, we run  $A(G, n)$ . This algorithm will return an independent set of cardinality at least  $f$  times the size of a maximum independent set in  $G$ .  $\square$

We define decision versions of the MWBIS and the MIVC problems as:-

$$\begin{aligned} \text{D-MWBIS} &= \{<G, w, k, t> \mid G \text{ is an undirected graph, } w : V(G) \rightarrow \mathbb{R}^+ \text{ is a} \\ &\quad \text{weight function and } G \text{ contains an independent set of cardinality} \\ &\quad \text{at most } k \text{ whose weight is at least } t\}. \\ \text{D-MIVC} &= \{<G, k, t> \mid G \text{ is an undirected graph and } G \text{ contains a set of} \\ &\quad \text{at most } k \text{ independent vertices that cover at least } t \text{ distinct} \\ &\quad \text{edges}\}. \end{aligned}$$

Since the MIS problem and the VC problem are known to be NP-hard, it is not surprising to see that the D-MIVC (and thereby the D-MWBIS) problem is NP-hard for general graphs. The following is an easy corollary to Theorem 1

**Corollary 2.** *D-MIVC problem is NP-hard.*

## 1.6 Motivation

The main motivation behind studying the MWBIS problem is the fact that it is a natural generalization of a very well-studied problem in the context of graph algorithms and approximation algorithms, the MWIS problem. MWIS problem finds application in wireless networks, scheduling, molecular biology, pattern recognition, coding theory, etc. MWBIS finds application in most scenarios where MWIS is used. Refer [5] to know about applications of MWBIS in ‘job scheduling in a computer’ and in ‘selecting non-interfering set of transmitters’.

## 1.7 Our contribution

We know that the MWIS problem on bipartite graphs is polynomial time solvable. Intuitively, the MWBIS problem, being a budgeted variant of the MWIS, is also expected to follow the same behaviour. But, contrary to our intuition, we show in Theorem 10 that the D-MIVC (and thereby the D-MWBIS) problem is NP-hard for bipartite graphs. This motivated us into looking at approximation algorithms. In Section 2.2, we give an  $O(nk)$  time,  $\frac{1}{2}$ -factor greedy approximation algorithm for the MWBIS problem on bipartite graphs. We give a tight example to the algorithm. In Section 2.3, we consider the LP relaxation of the integer program for MWBIS given in Section 1.4. We show that the integrality gap for this LP is upper bounded by  $\frac{1}{2} + \epsilon$  for bipartite graphs, where  $\epsilon$  is any number greater than 0. In other words, no LP-based technique, that uses this natural LP relaxation of MWBIS, is going to give us a better factor approximation algorithm for the MWBIS problem on bipartite graphs.

## 1.8 Notational note

Throughout the paper, we consider only finite, undirected, and simple graphs. For a graphs  $G$ , we shall use  $V(G)$  to denote its vertex set and  $E(G)$  to denote its edge set. For any  $S \subseteq V(G)$ , we shall use  $|S|$  to denote the cardinality of the set  $S$ . If a weight function  $w : V(G) \rightarrow \mathbb{R}^+$  is given, then  $\|S\|$  shall be used to denote the sum of the weights of the vertices in  $S$ . Otherwise,  $\|S\|$  will denote the number of edges in  $G$  having at least one endpoint in  $S$ . For any vertex  $v$  in the graph under consideration, we shall use  $\deg(v)$  to denote its degree in the graph. For any positive integer  $n$ , we shall use  $[n]$  to denote the set  $\{1, \dots, n\}$ .

## 2 MWBIS in bipartite graphs

We begin by first showing that D-MIVC on bipartite graphs is NP-hard in Section 2.1. We give a  $1/2$  factor approximation algorithm for MWBIS in Section 2.2 and show in Section 2.3 that LP-based methods are unlikely to yield better approximation algorithms.

### 2.1 NP-hardness of D-MIVC in bipartite graphs

We show that D-MIVC problem on bipartite graphs is NP-hard by reducing from  $(n-4, k)$ -CLIQUE problem. The  $(n-4, k)$ -CLIQUE problem is the famous clique problem on a graph, where the graph under consideration is an  $(n-4)$ -regular graph on  $n$  vertices.

$(n-4, k)$ -CLIQUE =  $\{ \langle G, k \rangle \mid G \text{ is an } (n-4)\text{-regular graph on } n \text{ vertices that contains a clique of size at least } k, \text{ where } k < \frac{n}{2} \}$ .

The  $(3, k)$ -Independent Set  $((3, k)\text{-IS})$  problem is about deciding whether there exists an independent set of size at least  $k$  or not in a given 3-regular graph. It was shown in [12] that the  $(3, k)$ -IS problem is NP-hard. This means that the  $(n-4, k)$ -CLIQUE problem is also NP-hard. Note that when  $k \geq \frac{n}{2}$ , the  $(3, k)$ -IS problem is about deciding whether the given graph is bipartite or not and is therefore polynomial-time solvable. Below we outline our reduction from the  $(n-4, k)$ -CLIQUE problem to D-MIVC.

**Construction 3.** Let  $n$  be a positive integer greater than 11 and let  $r = n - 4$ . Given an  $r$ -regular graph  $G$  with  $V(G) = \{v_1, \dots, v_n\}$  and  $E(G) = \{e_1, \dots, e_m\}$ , we construct a bipartite graph  $H$  with bipartition  $\{A, B\}$ , where  $A = \{a_1, \dots, a_m\}$ ,  $B = \beta \cup \Pi$  with  $\beta = \{b_1, \dots, b_n\}$  and  $\Pi = \{p_{i,j} \mid i \in [m], j \in [r-3]\}$ . The edge set  $E(H) = \{a_i p_{i,j} \mid i \in [m], j \in [r-3]\} \cup \{a_i b_j \mid \text{edge } e_i \text{ is incident on vertex } v_j \text{ in } G\}$ . Note that,  $\deg(a_i) = 2 + (r-3) = r-1$ ,  $\deg(b_i) = r$ , and every  $p_{i,j}$  is a pendant vertex.

**Observation 4.** In Construction 3, the subgraph of  $H$  induced on the vertex set  $A \cup \beta$  is isomorphic to the incidence graph of  $G$ . Thus,  $H$  is isomorphic to the incidence graph of  $G$  with  $r-3$  pendant vertices (, namely the  $p_{i,j}$  vertices) hanging down from each  $a_i$ .

In Lemma 5, we prove that if the graph  $G$  contains a clique of size  $k$ , then there exist some  $k+x$  independent vertices in  $H$  that cover at least  $kr+x(r-1)$  edges, where  $x = m - (kr - \binom{k}{2})$ . In Lemma 9, we prove the reverse implication of this statement.

**Lemma 5.** If there exists a clique of size  $k$  in the  $r$ -regular graph  $G$  given in Construction 3, then there exists  $k+x$  independent vertices in the bipartite graph  $H$  constructed that cover at least  $kr+x(r-1)$  edges, where  $x = m - (kr - \binom{k}{2})$ .

*Proof.* Without loss of generality, let us assume that the set of vertices  $S = \{v_1, \dots, v_k\}$  form a  $k$ -clique in  $G$ . It is easy to verify that  $\|S\| = kr - \binom{k}{2}$  and therefore the number of edges not incident on the vertices of  $S$  in  $G$  is  $x = m - \|S\| = m - (kr - \binom{k}{2})$ . Without loss of generality, let these  $x$  edges be  $e_1, \dots, e_x$ . Let  $T = \{a_1, \dots, a_x, b_1, \dots, b_k\}$ . Clearly,  $T$  is a set of independent vertices in  $H$  and  $\|T\| = kr + x(r-1)$ .  $\square$

Before we prove Lemma 9, we prove a few supporting lemmas, namely Lemmas 6 to 8, that give us some insight into the structure of  $G$  and  $H$ .

**Lemma 6.** Suppose there exists no clique of size  $k$  in the  $r$ -regular graph  $G$  given in Construction 3. Let  $i$  be an integer such that  $0 \leq i \leq r-k$ . Let  $p_{k+i}(G) = \max\{m - \|S\| : S \subseteq V(G), |S| = k+i\}$ . Then,

- (i)  $p_k(G) \leq x-1$ ,
- (ii) for every  $i \in [r-k]$ ,  $p_{k+i}(G) \leq p_{k+i-1}(G) - (r - (k+i-2))$ .

*Proof.* For every  $i$ , let  $S_{k+i}$  (, which is a subset of  $V(G)$ ,) be a set such that  $|S_{k+i}| = k+i$  and  $m - ||S_{k+i}|| = p_{k+i}(G)$ .

(i) Since  $G$  does not contain any  $k$ -clique,  $||S_k|| \geq kr - ((\binom{k}{2} - 1)$  and thus  $p_k(G) = m - ||S_k|| \leq m - (kr - ((\binom{k}{2} - 1)) = x - 1$ .

(ii) Let  $i \in [r-k]$ . Since  $G$  does not contain any  $k$ -clique,  $\exists v \in S_{k+i}$  such that  $v$  has at most  $k+i-2$  neighbours in  $S_{k+i}$  and therefore at least  $r - (k+i-2)$  neighbours outside  $S_{k+i}$ . Let  $X_{k+i-1} = S_{k+i} \setminus \{v\}$ . Then,  $||S_{k+i}|| \geq ||X_{k+i-1}|| + r - (k+i-2) \geq ||S_{k+i-1}|| + r - (k+i-2)$  (from the definition of  $S_{k+i-1}$ ). Therefore,  $p_{k+i}(G) = m - ||S_{k+i}|| \leq m - (||S_{k+i-1}|| + r - (k+i-2)) \leq p_{k+i-1}(G) - (r - (k+i-2))$ .  $\square$

Below, we introduce a couple of new definitions. Consider the bipartite graph  $H$  with bipartition  $\{A, B\}$  (, where  $B = \beta \uplus \Pi$ ,) constructed from  $G$  in Construction 3. Let  $i$  be an integer such that  $0 \leq i \leq r-k$ .

**Definition 4.** For any set  $S \subseteq \beta$ , let  $q(S) := \max\{|X| : X \subseteq A \text{ and no vertex in } X \text{ is adjacent with any vertex in } S\}$ . Then, we define  $q_{k+i}(H) := \max\{q(S) : S \text{ is a } (k+i)\text{-sized subset of } \beta\}$ .

**Definition 5.** Let  $\mathcal{I}_{k+i} = \{I \subseteq V(H) \mid I \text{ is a } (k+x)\text{-sized independent set, and } |I \cap \beta| = k+i\}$ , where  $x = m - (kr - \binom{k}{2})$ . Then, we define  $s_{k+i}(H) := \max\{|I| : I \in \mathcal{I}_{k+i}\}$ .

Suppose the graph  $G$  does not contain any  $k$ -clique. Then, from Observation 4, Lemma 6, and Definition 4, we have  $p_{k+i}(G) = q_{k+i}(H)$ . This gives us the following lemma.

**Lemma 7.** Suppose there exists no clique of size  $k$  in the  $r$ -regular graph  $G$  given in Construction 3. Consider the bipartite graph  $H$  constructed from  $G$ . We have,

- (i)  $q_k(H) \leq x - 1$ ,
- (ii) for every  $i \in [r-k]$ ,  $q_{k+i}(H) \leq q_{k+i-1}(H) - (r - (k+i-2))$ .

**Lemma 8.** Suppose there exists no clique of size  $k$  in the  $r$ -regular graph  $G$  given in Construction 3. Consider the bipartite graph  $H$  constructed from  $G$ . Then,  $s_r(H) < s_{r-1}(H) < \dots < s_k(H) < kr + x(r-1)$ , where  $x = m - (kr - \binom{k}{2})$ .

*Proof.* Let  $i$  be an integer such that  $0 \leq i \leq r-k$ . Let  $S_{k+i}$  be a  $(k+x)$ -sized independent set in  $H$  having exactly  $k+i$  vertices from the set  $\beta$  such that  $||S_{k+i}||$  is maximum among all such independent sets. That is,  $||S_{k+i}|| = s_{k+i}(H)$  (from the definition of  $s_{k+i}(H)$ ). Recall that  $S_{k+i} \subseteq V(H) = A \uplus \beta \uplus \Pi$  and from its definition we have  $|S_{k+i} \cap (A \uplus \Pi)| = x-i$ . Since the degree of any vertex in  $A$  (which is  $r-1$ ) is greater than the degree of any vertex in  $\Pi$  (which is 1), we have  $||S_{k+i}||$  maximized when  $S_{k+i} \cap A$  is maximized. Then, from the definition of  $S_{k+i}$  and by Definition 4, we get  $|S_{k+i} \cap A| = q_{k+i}(H)$ . Therefore,

$$\begin{aligned} s_{k+i}(H) &= ||S_{k+i}|| \\ &= ||S_{k+i} \cap \beta|| + ||S_{k+i} \cap A|| + ||S_{k+i} \cap \Pi|| \\ &= |S_{k+i} \cap \beta|r + |S_{k+i} \cap A|(r-1) + |S_{k+i} \cap \Pi| \\ &= (k+i)r + (q_{k+i}(H))(r-1) + (k+x) - (k+i + (q_{k+i}(H))). \end{aligned} \quad (2)$$

By Statement (i) of Lemma 7, we have  $q_k(H) = x - a$ , where  $a \geq 1$ . Therefore,  $s_k(H) = kr + (x-a)(r-1) + a \leq kr + (x-1)(r-1) + 1 < kr + x(r-1)$ .

Let  $i$  be an integer such that  $0 \leq i < r-k$ . Then,

$$\begin{aligned} s_{k+i}(H) - s_{k+i+1}(H) &= (q_{k+i}(H) - q_{k+i+1}(H))(r-2) - (r-1) \quad (\text{from Eqn. (2)}) \\ &\geq (r - (k+i-1))(r-2) - (r-1) \quad (\text{by (ii) in Lemma 7}) \\ &\geq 2(r-2) - (r-1) \quad (\text{since } i \leq r-k-1) \\ &\geq 5 \quad (\text{since } r = n-4 \text{ and } n > 11). \end{aligned}$$

□

**Lemma 9.** *If the  $r$ -regular graph  $G$  given in Construction 3 does not contain any clique of size  $k$ , then no set of  $k + x$  independent vertices in the bipartite graph  $H$  constructed covers  $kr + x(r - 1)$  edges or more, where  $x = m - (kr - \binom{k}{2})$ .*

*Proof.* Suppose the graph  $G$  given in Construction 3 does not contain any clique of size  $k$ . We shall then show that no set of  $k + x$  independent vertices in the bipartite graph  $H$  constructed from  $G$  covers  $kr + x(r - 1)$  edges or more

Let  $Y$  be an independent set of vertices of cardinality at most  $k + x$  in  $H$ . Then,  $Y = Y_A \uplus Y_\beta \uplus Y_\Pi$ , where  $Y_A = Y \cap A$ ,  $Y_\beta = Y \cap \beta$ , and  $Y_\Pi = Y \cap \Pi$ . The contribution of any vertex  $v \in Y$  to the coverage  $\|Y\|$  is its degree and this is highest when  $v \in Y_\beta$  (then,  $\deg(v) = r$ ) and lowest when  $v \in Y_\Pi$  (in this case,  $\deg(v) = 1$ ). When  $v \in Y_A$ ,  $\deg(v) = r - 1$ . Below, we shall prove that  $\|Y\| < kr + x(r - 1)$ . The proof is split into 2 cases based on the cardinality of  $Y_\beta$ .

*Case 1* ( $|Y_\beta| < k$ ).

Let  $|Y_\beta| = k - a$ , where  $a \geq 1$ . Then,  $\|Y\| = |Y_\beta|r + |Y_A|(r - 1) + |Y_\Pi| \leq (k - a)r + (x + a)(r - 1) \leq (k - 1)r + (x + 1)(r - 1) = kr + x(r - 1) - 1$ .

*Case 2* ( $|Y_\beta| \geq k$ ).

We split this case into two subcases:-

(i)  $|Y_\beta| \leq r$  ( $= n - 4$ ).

Let  $0 \leq i \leq r - k$ . Let  $|Y_\beta| = k + i$ . Then, by Definition 5 and Lemma 8,  $\|Y\| \leq s_{k+i} < kr + x(r - 1)$ .

(ii)  $|Y_\beta| > r$  ( $= n - 4$ ).

Since  $|\beta| = n$ , by definition of  $Y_\beta$ ,  $|Y_\beta| \leq n$ . Let  $i \in \{0, \dots, 3\}$ . Let  $|Y_\beta| = n - i$ . The set  $Y$  being an independent set in  $H$ , by Observation 4, the vertices in  $Y_A$  correspond to edges that are ‘outside’ the set of vertices in  $G$  that correspond to  $Y_\beta$ . As the cardinality of  $Y_\beta$  is close to  $n$  (in fact, at least  $n - 3$ ), we get  $|Y_A| \leq i$ . We are now prepared to estimate the coverage of  $Y$ .

$$\begin{aligned} \|Y\| &= |Y_\beta|r + |Y_A|(r - 1) + |Y_\Pi| \\ &\leq (n - i)r + i(r - 1) + ((k + x) - n) \\ &\leq nr + k + x - n \quad (\text{since } i = 0 \text{ yields the highest value}) \end{aligned}$$

In order to prove subcase (ii), we need to show that

$$\begin{aligned} kr + x(r - 1) &> nr + k + x - n \\ \text{i.e. } x(r - 2) &> (n - k)(r - 1) \quad (\text{rearranging terms}) \\ \text{i.e. } x &> (n - k) \left(1 + \frac{1}{r - 2}\right) \end{aligned}$$

Since  $r = n - 4$  and  $n > 11$ , we have  $r \geq 8$ . Thus, in order to prove this subcase it is enough to prove the following inequality:-

$$x > \frac{7}{6}(n - k).$$

We know that  $x = m - \left(kr - \binom{k}{2}\right) = \frac{nr}{2} - \left(kr - \frac{k^2 - k}{2}\right) = \frac{nr - 2kr + k^2 - k}{2} = \frac{r(n - k) - kr + k^2 - k}{2} = \frac{r(n - k) - k(n - 4) + k^2 - k}{2} = \frac{(r - k)(n - k) + 3k}{2} > \frac{((n - 4) - k)(n - k)}{2} \geq \frac{(n - 7)(n - k)}{4}$  (since  $k \leq \frac{n-1}{2}$ . See the definition of  $(n - 4, k)$ -CLIQUE in the beginning of this section.). Since  $n > 11$ , we get  $x \geq \frac{5}{4}(n - k) > \frac{7}{6}(n - k)$ .

□

Lemma 5 and Lemma 9 imply the following theorem.

**Theorem 10.** *D-MIVC problem on bipartite graphs is NP-hard.*

## 2.2 A $\frac{1}{2}$ -approximation algorithm

Consider the following greedy approximation algorithm for the MWBIS problem in bipartite graphs.

---

**Algorithm 1**  $\frac{1}{2}$ -approximation algorithm for the MWBIS problem on bipartite graphs

---

**INPUT:** A bipartite graph  $G$  with bipartition  $\{A, B\}$ , a weight function  $w : V(G) \rightarrow \mathbb{R}^+$ , and a positive integer  $k$ .

**OUTPUT:** An independent set of size at most  $k$ .

- 1: Let  $S_A$  denote the set of  $k$  highest weight vertices in  $A$ . In the case when  $k > |A|$ ,  $S_A = A$ . Similarly, let  $S_B$  denote the set of  $k$  highest weight vertices in  $B$ . In the case when  $k > |B|$ ,  $S_B = B$ . Find  $S_A$  and  $S_B$ .
- 2: If  $\|S_A\| \geq \|S_B\|$ , return  $S_A$ . Otherwise, return  $S_B$ .

---

We claim that Algorithm 1 is a  $\frac{1}{2}$ -factor approximation algorithm for MWBIS in bipartite graphs. Without loss of generality, let us assume that Algorithm 1 returns the set  $S_A$ . Let  $O$  be an optimal solution for the MWBIS problem on the bipartite graph  $G$  under consideration. Let  $O_A = O \cap A$  and  $O_B = O \cap B$ . Then  $\|O\| = \|O_A\| + \|O_B\| \leq \|S_A\| + \|S_B\| \leq 2\|S_A\|$ . Hence, we prove the claim. It is easy to see that Algorithm 1 runs in  $O(nk)$  time.

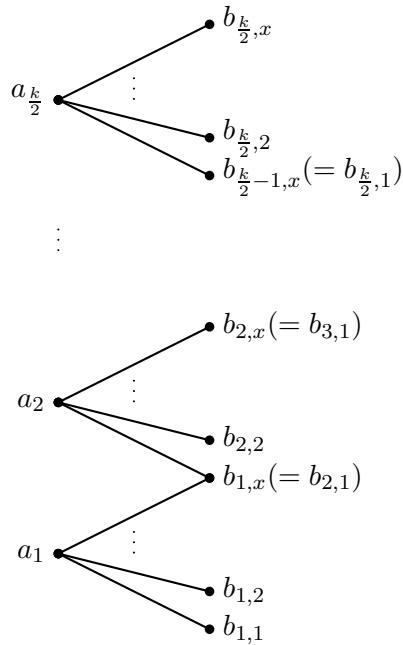


Figure 1: Graph  $H_1$

### 2.2.1 Tight example

Below we describe the construction of a bipartite graph  $H$  that is a tight example to our analysis of Algorithm 1. Let  $H_1$  be the connected, bipartite graph given in Figure 1. Let  $H_2$  be a graph isomorphic to  $H_1$  (one can imagine  $H_2$  to be a mirror copy of Figure 1 drawn right above it). We obtain the connected, bipartite graph  $H$  from the union of  $H_1$  and  $H_2$  by adding an edge connecting a pendant vertex in  $H_1$  with a pendant vertex in  $H_2$ . More formally,  $V(H) = V(H_1) \cup V(H_2)$  and  $E(H) = E(H_1) \cup E(H_2) \cup \{h_1, h_2\}$ , where  $h_1$  and  $h_2$  are any two pendant vertices in  $H_1$  and  $H_2$ , respectively. We claim that  $H$  is a tight example for Algorithm 1. Consider the MIVC problem on graph  $H$  with parameter  $k$ . The optimal solution chooses all the degree  $x$  vertices. This gives a solution of size  $kx$ . Since each part in  $H$  has exactly  $\frac{k}{2}$  vertices of degree  $x$  and the remaining vertices are each of degree at most 2, our algorithm returns a solution of size at most  $\frac{kx}{2} + \frac{k}{2}(2) = \frac{kx}{2} + k$ . Thus, the approximation ratio is at most  $\frac{\frac{kx}{2} + k}{kx} = \frac{1}{2} + \frac{1}{x}$ . This ratio approaches  $\frac{1}{2}$  as  $x$  tends to infinity.

### 2.2.2 Extending Algorithm 1 to general graphs

Algorithm 1 can be generalized to an arbitrary graph  $G$  (that is not necessarily bipartite) provided we are given a proper vertex coloring of  $G$ . Suppose we are given a proper coloring of the vertices of  $G$  using  $p$  colors that partitions the vertices into  $p$  color classes namely  $C_1, \dots, C_p$ . Let  $S_1, \dots, S_p$  denote sets of  $k$  highest degree vertices in color classes  $C_1, \dots, C_p$ , respectively. For any  $S_i$ , if  $k > |C_i|$ , then  $S_i = C_i$ . Find a set  $S_i$  such that  $\|S_i\| \geq \|S_j\|, \forall j \in [p]$ . Return  $S_i$ .

This algorithm gives us a  $\frac{1}{p}$ -approximation for MWBIS. Suppose it returns a set  $S_j$  as the output. Let  $O$  be an optimal solution for the MWBIS problem on  $G$ . Let  $O_i = O \cap C_i$  for  $i \in [p]$ . Then  $\|O\| = \|O_1\| + \dots + \|O_p\| \leq p\|S_j\|$ . It is easy to see that the algorithm runs in  $O(nk)$  time.

We know that we can properly color any graph  $G$  with  $\Delta + 1$  colors using a greedy coloring algorithm, where  $\Delta$  denotes the maximum degree of  $G$ . This gives us a  $\frac{1}{\Delta+1}$ -approximation algorithm for MWBIS on  $G$ . Since  $d$ -degenerate graphs can be properly colored using  $d + 1$  colors in linear time, this algorithm is a  $\frac{1}{d+1}$ -factor approximation algorithm for such graphs.

## 2.3 LP relaxation and a matching integrality gap

Let us take another look at Integer Program (1) given in Section 1.4 for the MWBIS problem on  $G$  (, where  $G$  is a graph with  $V(G) = \{v_1, \dots, v_n\}$  and  $w : V(G) \rightarrow \mathbb{R}^+$ ). Recall, we use  $\mathcal{C}$  to denote the set of all maximal cliques in  $G$ . We obtain the LP relaxation of this program by changing the domain of variable  $x_i$  from  $x_i \in \{0, 1\}$  to  $x_i \geq 0$ .

$$\begin{aligned}
 & \text{Maximize} \sum_{i \in [n]} w(v_i) \cdot x_i & (3) \\
 & \text{s.t.} \sum_{i=1}^n x_i \leq k \\
 & \quad \sum_{i: v_i \in C} x_i \leq 1, \quad \forall C \in \mathcal{C} \\
 & \quad x_i \geq 0, \quad \forall i \in [n].
 \end{aligned}$$

We now present an instance of the MIVC problem to illustrate that the integrality gap of the above LP is upper bounded by  $\frac{1}{2} + \epsilon$  (where  $\epsilon$  is any number greater than 0) for bipartite graphs.

**Example 11.** Let  $H$  be a bipartite graph with bipartition  $\{A, B\}$ , where  $|B| = k(k-1)+1 = p$  and  $|A| = (k-1)p+1$ . Let  $A = \{a_{i,j} \mid i \in [p], j \in [k-1]\} \cup \{a_0\}$  and  $B = \{b_1, \dots, b_p\}$ . The edge set  $E = \{a_0b_i \mid i \in [p]\} \cup \{a_{i,j}b_i \mid i \in [p], j \in [k-1]\}$ . Note that  $\deg(a_0) = p$ ,  $\deg(b_i) = k$ , and  $\deg(a_{i,j}) = 1$ . For every  $v \in V(H)$ , let  $w(v) = \deg(v)$ . Figure 2 illustrates a drawing of  $H$  with  $k = 3$ .

The best integral solution to the above LP for  $H$  is to select vertex  $a_0$  and  $k-1$  vertices of degree 1 from the set  $\{a_{i,j} \mid i \in [p], j \in [k-1]\}$ . The coverage by these  $k$  vertices is  $\deg(a_0) + (k-1) = p + (k-1) = k(k-1) + 1 + (k-1) = k^2$ .

The  $x$ -variables in the LP have the following associations:  $x_0$  is associated with vertex  $a_0$ ,  $x_i$  with  $b_i$ , and  $x_{i,j}$  with  $a_{i,j}$ . One possible fractional solution to the above LP for  $H$  is to assign  $x_0 = \frac{k-1}{k}$ ,  $x_i = \frac{1}{k}$ ,  $\forall i \in [p]$ , and  $x_{i,j} = 0$ ,  $\forall i \in [p], \forall j \in [k-1]$ . This is a feasible solution to the above LP as  $x_0 + \sum_{i \in [p], j \in [k-1]} x_{i,j} + \sum_{i \in [p]} x_i = \frac{k-1}{k} + 0 + p \frac{1}{k} = k$  and sum of the  $x$ -values of every maximal clique is at most 1. The number of edges covered by this fractional solution is  $(\frac{k-1}{k} + \frac{1}{k})p + (\frac{1}{k}p(k-1)) = p(1 + \frac{k-1}{k}) = (k(k-1) + 1)(\frac{2k-1}{k})$ . Hence the upper bound for integrality gap of the above LP established by this example is  $\frac{k^2}{(k(k-1)+1)(\frac{2k-1}{k})} = \frac{k^3}{(k^2-k+1)(2k-1)} = \frac{k^3}{2k^3-3k^2+3k-1} = \frac{1}{2-\frac{3}{k}+\frac{3}{k^2}-\frac{1}{k^3}} = \frac{1}{2-\frac{1}{k}(3+\frac{1}{k^2}-\frac{3}{k})}$ . This value approaches  $1/2$  as  $k$  tends to infinity. Thus, the integrality gap of the above LP is upper bounded by  $\frac{1}{2} + \epsilon$  (where  $\epsilon$  is any number greater than 0) for bipartite graphs.

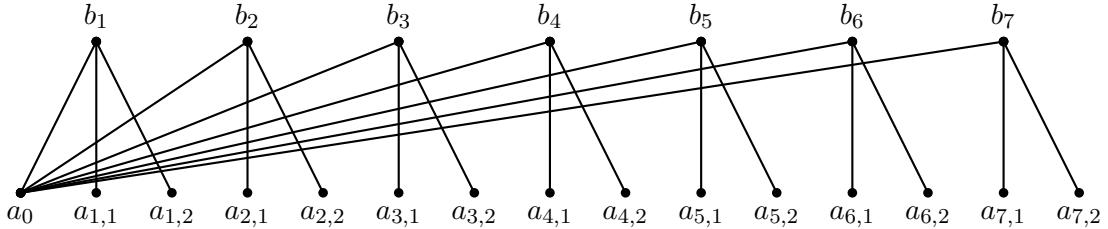


Figure 2: A drawing of the bipartite graph  $H$  described in Example 11 with  $k = 3$ .

### 3 Concluding remarks

In this paper we study the MWBIS problem, a generalization of MWIS, mainly in the context of bipartite graphs. In Theorem 10, we prove that MWBIS on bipartite graphs is NP-hard and then we propose an easy, greedy  $\frac{1}{2}$ -approximation algorithm, whose factor of approximation matches with the integrality gap of a natural LP relaxation of the problem. It would be interesting to see if one can improve the factor of approximation using a more clever and sophisticated approach.

It is well-known that the MWIS problem in general graphs does not have a polynomial time constant factor approximation algorithm, unless  $P=NP$ . Håstad [15] showed that there is no  $\frac{1}{n^{1-\epsilon}}$ -factor approximation algorithm for MWIS, where  $\epsilon > 0$ , assuming NP-hard problems have no randomized polynomial time algorithms. For every sufficiently large  $\Delta$ , there is no  $\Omega(\frac{\log^2 \Delta}{\Delta})$ -factor polynomial time approximation algorithm for MWIS in a degree- $\Delta$  bounded graph, assuming the unique games conjecture and  $P \neq NP$  [4]. As

MWBIS is a generalization of MWIS, these results hold true for the MWBIS problem too. Let  $G$  be a graph on  $n$  vertices with the degree of any vertex being at most  $\Delta$ . Now, suppose we have a  $t$ -factor approximation algorithm  $A(G)$  for finding MWIS in  $G$ . Then, we can obtain a  $\frac{kt}{n}$ -factor approximation algorithm for MWBIS in  $G$ . Let  $A(G)$  return an independent set  $S$  as output. Let  $S'$  denote the set of  $k$  highest weight vertices in  $S$ . If  $k > |S|$ , then  $S' = S$ . Return  $S'$ . Let  $O, O'$  be optimal solutions for the MWIS and the MWBIS problem in  $G$ , respectively. We have  $\|S'\| \geq \frac{k\|S\|}{|S|} \geq \frac{k\|S\|}{n} \geq \frac{kt\|O\|}{n} \geq \frac{kt\|O'\|}{n}$ . Boppana and Halldórsson in [8] gave an  $\Omega(\log^2 n/n)$ -factor approximation algorithm for MWIS. This yields an  $\Omega(k \log^2 n/n^2)$ -factor algorithm for MWBIS. For bounded degree graphs, Halldórsson in [13] gave an  $\Omega(\sqrt{\log \Delta}/\Delta)$ -factor approximation algorithm for MWIS. This gives an  $\Omega(k\sqrt{\log \Delta}/n\Delta)$ -factor approximation algorithm for MWBIS. Finding better approximation algorithms for MWBIS in general graphs is another nice open question to pursue.

## References

- [1] Alexander A Ageev and Maxim I Sviridenko. Approximation algorithms for maximum coverage and max cut with given sizes of parts. In *Integer Programming and Combinatorial Optimization*, pages 17–30. Springer, 1999.
- [2] Nicola Apollonio and Bruno Simeone. Improved approximation of maximum vertex coverage problem on bipartite graphs. *SIAM Journal on Discrete Mathematics*, 28(3):1137–1151, 2014.
- [3] Nicola Apollonio and Bruno Simeone. The maximum vertex coverage problem on bipartite graphs. *Discrete Applied Mathematics*, 165:37–48, 2014.
- [4] Per Austrin, Subhash Khot, and Muli Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. In *Computational Complexity, 2009. CCC’09. 24th Annual IEEE Conference on*, pages 74–80. IEEE, 2009.
- [5] Sayan Bandyapadhyay. A variant of the maximum weight independent set problem. *CoRR*, abs/1409.0173, 2014.
- [6] Reuven Bar-Yehuda. Using homogeneous weights for approximating the partial cover problem. *Journal of Algorithms*, 39(2):137–144, 2001.
- [7] Piotr Berman and Marek Karpinski. On some tighter inapproximability results, further improvements. 1998.
- [8] Ravi Boppana and Magnús M Halldórsson. Approximating maximum independent sets by excluding subgraphs. *BIT Numerical Mathematics*, 32(2):180–196, 1992.
- [9] Nader H Bshouty and Lynn Burroughs. Massaging a linear programming solution to give a 2-approximation for a generalization of the vertex cover problem. In *STACS 98*, pages 298–308. Springer, 1998.
- [10] Bugra Caskurlu, Vahan Mkrtchyan, Ojas Parekh, and K Subramani. On partial vertex cover and budgeted maximum coverage problems in bipartite graphs. In *Theoretical Computer Science*, pages 13–26. Springer, 2014.
- [11] Rajiv Gandhi, Samir Khuller, and Aravind Srinivasan. Approximation algorithms for partial covering problems. *Journal of Algorithms*, 53(1):55–84, 2004.

- [12] Michael R Garey, David S. Johnson, and Larry Stockmeyer. Some simplified np-complete graph problems. *Theoretical computer science*, 1(3):237–267, 1976.
- [13] Magnús M Halldórsson. Approximations of weighted independent set and hereditary subset problems. *Journal of Graph Algorithms and Applications*, 4(1):1–16, 2000.
- [14] Qiaoming Han, Yinyu Ye, Hantao Zhang, and Jiawei Zhang. On approximation of max-vertex-cover. *European Journal of Operational Research*, 143(2):342–355, 2002.
- [15] Johan Håstad. Clique is hard to approximate within  $n^{1-\epsilon}$ . In *Foundations of Computer Science, 1996. Proceedings.*, 37th Annual Symposium on, pages 627–636. IEEE, 1996.